The temporal evolution of a far-right forum

Kleinberg, Bennett; van der Vegt, Isabelle; Gill, Paul

Published in:
Journal of Computational Social Science

DOI:
10.1007/s42001-020-00064-x

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Journal of Computational Social Science
https://doi.org/10.1007/s42001-020-00064-x

RESEARCH ARTICLE

The temporal evolution of a far-right forum

Bennett Kleinberg[1,2] · **Isabelle van der Vegt**[2] · **Paul Gill**[2]

Received: 26 October 2019 / Accepted: 3 February 2020

Abstract

The increased threat of right-wing extremist violence necessitates a better understanding of online extremism. Radical message boards, small-scale social media platforms, and other internet fringes have been reported to fuel hatred. The current paper examines data from the right-wing forum Stormfront between 2001 and 2015. We specifically aim to understand the development of user activity and the use of extremist language. Various time-series models depict posting frequency and the prevalence and intensity of extremist language. Individual user analyses examine whether some super users dominate the forum. The results suggest that structural break models capture the forum evolution better than stationary or linear change models. We observed an increase of forum engagement followed by a decrease towards the end of the time range. However, the proportion of extremist language on the forum increased in a step-wise matter until the early summer of 2011, followed by a decrease. This temporal development suggests that forum rhetoric did not necessarily become more extreme over time. Individual user analysis revealed that super forum users accounted for the vast majority of posts and of extremist language. These users differed from normal users in their evolution of forum engagement.

Keywords Extremist language · Right-wing extremism · Forum data · Linguistic analysis · Online engagement

Introduction

In 2008, a user on the right-wing extremist forum Stormfront argued that Britain was to be faced by a civil war due to Muslim immigration. A few years later, on 22 July 2011, the same user, later identified as Anders Breivik, killed 77 people in attacks in the centre of Oslo and Utøya island, Norway. Shortly before these attacks, he had posted a 1500-page manifesto on the forum, in addition to emailing it to thousands

✉ Bennett Kleinberg
bennett.kleinberg@ucl.ac.uk

[1] Dawes Centre for Future Crime, University College London, London, UK

[2] Department of Security and Crime Science, University College London, London, UK

Published online: 21 February 2020

of people. This incident does not stand in isolation. Recent terrorist attackers are also known to have been active users of extremist forums and social media platforms. This includes the Christchurch attacker in 2018, whose actions inspired other acts of violence including tragedies in El Paso, Texas and Oslo, Norway [1].

Stormfront.org was one of the early online extremist discussion forums, launched in 1995 by a white nationalist and former Ku Klux Klan leader. Throughout the years, Stormfront has become a breeding ground for right-wing extremists worldwide [2]. It has been reported that a disproportionate number of Stormfront users have been responsible for hate crimes and (mass) murders ever since the site's founding [3]. New extremist online spaces are constantly springing up, including online discussion boards such as 8Kun (formerly 8Chan [4]), 4Chan [5], and Gab [6]. It has been reported that "alternative social media platforms, image boards, fringe forums and encrypted chat channels are instrumental in diffusing influential ideologies that propagate hatred and violence" [7]. While government and industry intensified efforts to tackle this problem (e.g., by means of the Global Internet Forum to Counter Terrorism; see [8], and the Christchurch call; see [9]), detecting and consequently removing extremist and hate-fuelled material remains an ongoing problem.

While the majority of previous research on extremist language focused on its detection, much less attention focuses upon understanding how the use of extreme language develops over time. Some preliminary findings suggest that external events such as elections or controversial decisions might affect language [10], while other studies did not find support for such a "ripple effect" [11]. An aspect that has received limited attention is how the language used on a niche forum evolves over a prolonged period. Understanding the temporal trajectory of language can shed light on profound questions such as whether communication is becoming more extreme over time or whether the prevalence and intensity of extremist language remain constant. Moreover, such an analysis can also illuminate the nature of language change; that is, whether language becomes continuously more (less) extreme over time, or whether there are sudden (phase) transitions to different discrete levels of extremist language.

For policymakers overseeing or providing a platform for a free exchange of ideas (e.g., Twitter, Reddit, Facebook), such a level of understanding could help to find points of intervention. For example, among the core challenges of social media companies today is the dilemma between allowing free speech and prohibiting overly aggressive, extremist or false content [12]. Arguably, identifying shifts in the use of aggressive language could inform policymakers about when to intervene and when to allow the discourse to happen. Likewise, law enforcement agencies, worried about an increased attraction of Internet users to extremist forums, could utilise an analysis of the temporal development of user activity to understand the degree to which user behaviour changes.

This paper investigates the temporal development of user activity and the use of extremist language on the white nationalist forum Stormfront.org. We examine the evolution of the forum for a time span of 14 years by assessing both forum engagement and the actual language of the forum posts. Specifically, we look at the number of posts and the length of the posts as proxies of user activity. For the content-based

analysis, we examine extremist language as both a binary (i.e., whether or not a post contains extremist language) and a continuous phenomenon (i.e., the intensity of extremist language). Our overall aim is to quantitatively test different possible temporal trajectories of extremist language. In focusing on the understanding of the evolution of language use and user activity, this paper seeks to add a dimension of analysis that is currently under-developed and might offer insights to researchers and policymakers beyond the mere identification of potentially worrying user accounts. Furthermore, this examination of a large time span of (early) extremist online activity may help to better understand emerging radical platforms, and develop mitigation strategies for these phenomena.

Background

The following section is split into three sub-sections. Since a core focus of this paper is to measure hateful language, the first sub-section outlines various attempts to automatically identify hate speech and its analogues. Second, since the paper additionally measures changes in sentiment over time, we discuss previous work within the context of extremism and radicalisation that has endeavoured to do the same. Lastly, we discuss studies of engagement amongst extremist communities that go beyond measuring hate speech and sentiment alone.

Identifying hateful language

Within the computational research literature, various studies attempt to automatically identify hate speech. Typically, approaches relied on linguistic markers of hate speech, such as the presence of derogatory terms. Siegel et al. [10] studied hate speech on Twitter during and after the 2016 American elections, focusing on tweets that mentioned Donald Trump or Hillary Clinton. They examined whether there was an increase in hate speech and white nationalist rhetoric resulting from Donald Trump's election campaign. They measured hate speech through word lists drawn from Hatebase, the Racial Slur Database (http://www.rsdb.org/), and the Anti-Defamation League's database of white nationalist language (https://www.adl.org/hate-symbols). They performed an interrupted time-series analysis which showed a spike in hate speech following the 2017 travel ban imposed by President Trump. However, there was no lasting increase in hate speech or white nationalist rhetoric observed in the data.

Moving beyond major social media outlets, other studies examined language within discussion boards of specific extremist groups. For example, [13] distinguished posts written by Stormfront users from lone-actor terrorist writings. They extracted linguistic features using linguistic inquiry and word count (LIWC). The LIWC software produces proportions of words in a text thought to reflect certain psycho-social constructs, such as positive emotion and power, as well as specific linguistic features, such as personal pronouns and auxiliary verbs. Using an Adaboost classification algorithm, this resulted in an overall accuracy of 0.90, recall of 0.93,

and precision of 0.93, with 'time', 'articles', 'personal pronouns', 'see', 'differentiation', 'prepositions', 'quantifiers', 'negative emotion,' 'biological processes', and 'cognitive processes' as the most essential features.

Shrestha et al. [14] used machine learning approaches to identify 'extreme adopters' within a Swedish xenophobic forum. They considered extreme adopters as forum users who strongly identify with the community and (partially) express this using forum-specific jargon. They determined jargon by manually examining a sample of 500 posts from which they extracted 150 unusual words or phrases (i.e., manually coded as uncommon in everyday language). They defined extreme adopters as those who used more than 30 jargon words with a term frequency above 0.005 ($n = 587$ users). In a supervised classification task (extreme adopters vs. other forum users), the 200 most frequent unigrams, bigrams, and pairs of letters (i.e., data-dependent features) were used as features, in addition to word length, letters, digits, punctuation, and LIWC categories (i.e., data-independent features). The high classification accuracies (80–86%) achieved with both types of features suggest that extreme adopters differ from other forum users in terms of jargon and other language features, some of which can be considered racist slurs.

Sentiment analysis

An alternative line of research focuses on measuring sentiment (i.e., positive and negative language) and affect (i.e., expressions of emotion) within and across hate forums. For example, hand-crafted specific lexicons for hate and violence were employed to measure affect intensities on American and Middle Eastern extremist group forums from the dark web [15]. Middle Eastern and US extremist forums contained similar levels of hate affect, but Middle Eastern forums scored higher in terms of violence intensity. Similarly, other researchers [16] measured aggression, racism and worries within three sub-forums of the Stormfront discussion board. Domain experts annotated a sample of 300 posts for their affect intensities (high or low). A supervised classification task used both the 100 most frequent words in the posts rated as high intensity and the 100 words that differed the most between high- and low-intensity post as features. In addition, they used all LIWC 2015 categories, word count, part-of-speech tags, and three 'expert knowledge' dictionaries for worries, racism, and aggression. Essential features for classification included the words 'black', 'race', the word stem 'immigr', and LIWC categories for religion and anger. Classification accuracies with different classifiers ranged between 80 and 93%, with recall rates of 61–76%.

Scrivens et al. [17] identified radical users across four Islamic web forums through sentiment. They tagged a sample of approximately one million posts for parts-of-speech and then created a list of keywords with the 100 most frequent nouns for each forum. They extracted all keywords from the forum posts and analysed them for positive or negative sentiment, resulting in an average sentiment score for each post. They then computed a composite 'radical score' for each forum user by summing scores for (1) the average sentiment across all posts by the user, (2) the volume of negative posts, (3) the severity of negative posts, and (4) the duration

of negative posts. By doing so, they found that the most radical users were concentrated on two of the four examined forums. The same type of 'radical score' was used to study a Canadian right-wing forum [18]. Specifically, trajectories of anti-Semitic, anti-black, and anti-LGBTQ posting behaviours over time were examined. It was found that the sample as a whole exhibited a steady increase of radical score over time, suggesting increased polarisation [18]. Forum users' individual radical posting behaviour, however, decreased within their first two years as a member.

Another study that has examined radicalisation trajectories based on sentiment includes an examination of six Islamic dark web forums [19]. Posts were assigned a sentiment score, and then aggregated for each month (across 12 years), which showed average neutral to negative sentiment. Specific attention was paid to terrorist attacks within the time span, with some 'spikes' of negative sentiment coinciding with significant events (e.g., the 2005 London bombings).

Temporal changes in sentiment have also been studied in ISIS magazines *Dabiq* and *Rumiyah* [20] between 2014 and 2017. Although language use in general was found to be relatively stable (negative) throughout the years, enemies were referred to with increasing negative sentiment over time. References to ISIS itself became more positive over time [20].

Engagement in extremism

In addition to measuring hateful language and sentiment, some work has also examined engagement amongst extremist communities, such as post sharing [21], comment activity [22], and link sharing [5]. Elsewhere, researchers examined a sample of 154K Twitter users and the extent to which they shared pro-ISIS content, defined as tweets from known pro-ISIS accounts or other accounts that were suspended for supporting ISIS [21]. Examining these patterns over time, the authors suggest most users became 'activated' in the summer of 2014 when ISIS shared many beheading videos. Furthermore, they investigated Twitter users' adoption of pro-ISIS language by examining tweets for pro-ISIS and anti-Western terms represented in a custom lexicon. They found 208 users in the sample who used pro-ISIS terms (e.g., "apostate", "caliphate", "ummah") more than five times, and used such terms more than anti-ISIS terms (e.g., "Daesh", a pejorative term for ISIS).

Ribeiro et al. [22] empirically examined whether YouTube provides a "radicalisation pipeline" by analysing contrarian videos, meaning that the creators of the videos strongly oppose mainstream views. They distinguished between videos posted by the 'alt-lite' (contrarians who deny embracing white nationalism), 'alt-right' (openly declared white nationalists) and the 'intellectual dark web' (IDW; a loosely defined group of iconoclastic thinkers, academics and media personalities). They examined the comment activity on videos and found extreme content had higher engagement levels. They additionally tracked commenters and found a percentage of users moved towards more extreme content throughout the years. For example, of users who commented on IDW and alt-lite videos between 2006 and 2012, about 10% commented on 'light' alt-right videos in 2018, and 4% commented on 'mild to severe' alt-right videos. Based on a simulation of recommendation algorithms, they

also found that alt-lite channels were easily found through IDW channels, and alt-right channels can also be found through the former two communities. These findings only held for channel recommendations, but not for video recommendations. In short, they concluded that YouTube indeed may provide a 'radicalisation pipeline'.

Studies have also focussed on user activity on other messaging boards such as 4chan [5] and Gab [6]. In an analysis of eight million posts on 4chan's /pol ('politically incorrect') board, it was found that the majority of outward links directed to YouTube and right-wing news outlets [5]. When examining the /pol board and comments on the YouTube videos that were linked to, a large number of hateful terms from the Hatebase directory were found. A study on the 'fringe community' Gab [6] revealed that most discussions on the platform centre around news, world events, and politics. For instance, activity increased leading up to the inauguration of Donald Trump, with further activity volatility coinciding with other meaningful events, such as the firing of the ex-FBI director James Comey, and the Charlottesville 'Unite the Right' rally. Additionally, 5.4% of the posts were found to contain hate words [4].

Aims of this study

This paper has three aims. First, we examine the temporal evolution of the user base's forum engagement using two proxy measures of forum engagement: posts per month and average post length. Second, we apply the same analytical approach to the content of the posts by examining the amount of posts with extremist language and the intensity of extremist language. Third, we investigate the user base more specifically in testing whether some users dominate the forum and how they might differ from "normal" forum users in their progression from starting on the forum towards the end of the dataset.

Method

The code and resources to reproduce the analyses reported in this paper are available at https://osf.io/eq7z6/. The dataset stems from a publicly accessible website (i.e. anyone with an Internet connection can access the website and obtain the data from it) and is available upon request.

Data

The dataset contained all posts made in the right-wing online forum Stormfront between September 2001 and February 2015. These data include the text of each post that was not a quote of another post as well as metadata such as the username and date. From the total number of posts ($n = 2{,}033{,}706$), we excluded those not written in English (12.35%), did not contain a username (5.59%) or were shorter than 15 words after the removal of stop words (32.33%). We also excluded all posts

made before the first of December, 2001 due to minimal forum activity (0.16%). The final sample consisted of 1,009,986 posts.

Modelling extremist language

For each post in the final dataset, we modelled the use of extremist language through two lexicons as proxy measures for profane language and racial slurs, as well as the sentiment of each post.

Profane language: We constructed a sub-dictionary for profane language from three lexicons contained in the R *lexicon* package [23]—the bannedwordlist.com repository (e.g., "bitch", "cock", $n=77$ words), Alejandro Alvarez's list of bad words (e.g., "n*gga", "fag*", $n=438$), and a Stackoverflow user's list of bad words (e.g., "negro", "cunt", $n=343$). We added a list of words banned by Google (e.g., "a55", "whore", $n=451$), Luis von Ahn's list of bad words (e.g., "zigabo", "pi55", $n=1383$), as well as the base lexicon of abusive language from Wiegand et al. [24] (e.g., "niglet", "latrino", $n=551$). After removing duplicates, we retained a profane language lexicon of 1572 words.

Racial slurs: We also created a lexicon of racial slurs from the Racial Slur Database (http://www.rsdb.org/) resulting in 2586 entries of racial slurs (e.g., "beekeeper", "bara").

Sentiment extraction: Finally, we measured each post's sentiment using a sentiment look-up table in an algorithm that considered valence shifters (e.g., "hardly", "not") of the sentiment values. Valence shifters can change the polarity of a sentiment ("don't like" vs "like") or amplify ("really bad" vs "bad") and de-amplify ("barely exciting" vs "exciting") a sentiment [25, 26]. Specifically, we built a context window of two words before and after a sentiment match and corrected the sentiment if valence shifters were present in the resulting 5-word context window. This approach does not rely on punctuation and is, therefore, suited for poorly or non-punctuated text data. Higher sentiment values indicate a post's more positive tone.

Binary and continuous measures of extremist language

We operationalized extremist language as a binary (i.e. extreme language vs non-extremist language) and a continuous construct (i.e. the intensity of extremist language). We labelled a forum post as "extremist" if it contained at least one racial slur, one profane word, and had an overall negative sentiment. Each post was, therefore, labelled as either "extremist" or "non-extremist". We measured a post's intensity through the sum of racial slurs and profane words minus the post's sentiment.

Analysis plan

The primary objective is to assess the model fit of various time-series models on the aggregated data by month. For the different outcome measures (forum engagement and extremist language), we first test whether the monthly aggregated time series is stationary through the Augmented Dickey–Fuller test. A stationarity time series

Table 1 Descriptive statistics of the final sample

	Mean	SD	Median	Range
Post length	77.46	152.13	42.00	16; 10,388
Profane language (%)	5.54	5.25	4.55	0.00; 87.50
Racial slurs (%)	2.38	2.91	1.67	0.00; 81.82
Sentiment	0.06	0.30	0.05	−2.25; 2.53
Number of posts per month	6392	2500	6947	1010; 12,376
Post length per month	77.16	8.07	77.08	60.37; 97.48
Extremist posts per month (%)	22.16	3.16	22.65	12.46; 27.79
Extremist language per month	5.71	0.75	5.72	3.60; 7.14

means that the process underlying the data does not change over time (e.g., a strictly seasonal pattern) but does not imply that the actual values of the time series remain constant. This procedure tests whether we can reject the null hypothesis of non-stationarity (i.e. a non-significant p value suggests that the data are not stationary [27]). We fit three models to the data and then compare fit indices: (i) a stationary, intercept-only model, (ii) a linear temporal trend model, and (iii) a breakpoint model. The intercept-only model fits a straight, horizontal line to the time series, assuming that the values remain constant over time. The linear change model regresses the observed values on the temporal progression in a strictly linear manner. The structural breakpoint model was calculated with the *strucchange* R package [28] and used the Bayesian information criterion (BIC) to find the optimal number, if any, of structural breaks in the data. These breakpoints were then fitted to the data, and the dated structural changes were obtained [29].

Since the models are not nested, we use the AIC [30], BIC, as well as the mean absolute error (MAE) and the root mean squared error (RMSE) as model evaluation indices for model comparisons.

Results

Descriptive statistics

The final sample contained 1,009,986 posts made by 41,741 self-identified unique users. The monthly aggregation resulted in data for 158 months from Jan. 2002 until Feb. 2015 (Table 1).

Forum-level analysis

Forum activity: For the number of posts per month, the Augmented Dickey–Fuller (ADF) test did not allow for the rejection of the non-stationarity null hypothesis,

Table 2 Model fit indices of the stationary, linear and breakpoint model for each of the four forum-level outcome measures

Outcome measure	Model	Evaluation metric			
		AIC	BIC	MAE	RMSE
Forum activity (number of posts)	Stationary model	2923.86	2929.99	1916.17	2492.72
	Linear model	2872.21	2881.40	1745.13	2103.49
	Breakpoint model (4)	**2669.76**	**2688.13**	**838.49**	**1087.54**
Number of words per post	Stationary model	1111.20	1117.33	6.72	8.04
	Linear model	1080.83	1090.01	5.82	7.26
	Breakpoint model (3)	**987.26**	**1002.57**	**4.10**	**5.33**
Proportion posts with extremist language	Stationary model	−640.15	−634.02	0.0237	0.0315
	Linear model	−731.20	−722.01	0.0194	0.0234
	Breakpoint model (4)	**−935.28**	**−916.90**	**0.0095**	**0.0121**
Average extremist language score per post	Stationary model	361.49	367.62	0.6343	0.7500
	Linear model	363.42	372.61	0.6332	0.7499
	Breakpoint model (2)	**196.44**	**208.69**	**0.3313**	**0.4393**

The integer in brackets after the breakpoint model indicates the number of regime shifts in the model. The best model fit is highlighted in bold

AIC Akaike's information criterion, *BIC* Bayesian information criterion, *MEA* mean absolute error, *RMSE* root mean square error

Dickey–Fuller test statistic = −1.29, $p = 0.872$.[1] The evaluation metrics in Table 2 indicate that a structural breakpoint model fits the time series of the number of posts per months best. Figure 1 shows that the breakpoint model contains four regime shifts with breaks at 11/2003, 10/2005, 10/2007 and 09/2009. The first three regimes represent a stepwise escalation of the number of posts until 9408 posts per months in 09/2009 from where the forum activity drops to 6890 posts per months. Overall, the forum activity analysis suggests that activity on the forum increased until it reached peak activity for almost 2 years between 10/2007 and 9/2009 and then faded off.

Post length

The ADF tests did not allow for the null hypothesis of non-stationarity, Dickey–Fuller statistic = −2.37, $p = 0.423$. The breakpoint model fitted the data best with structural breaks at 3/2004, 11/2006 and 5/2011 (Table 2). The average post length was highest in the regime between 3/2004 and 11/2006 with 85.04 words per post and decreased afterwards. The lowest level was reached after 5/2011 until the end of the data, with 68.93 words per post (Fig. 2). These findings indicate that after the regime increase in 3/2004, the length of the posts consistently declined and reached the lowest levels after 5/2011.

[1] All ADF tests were calculated with the default lag-order = 5 of the *tseries* R package. The conclusions about the rejection of the null hypothesis do were not affected by the lag-orders.

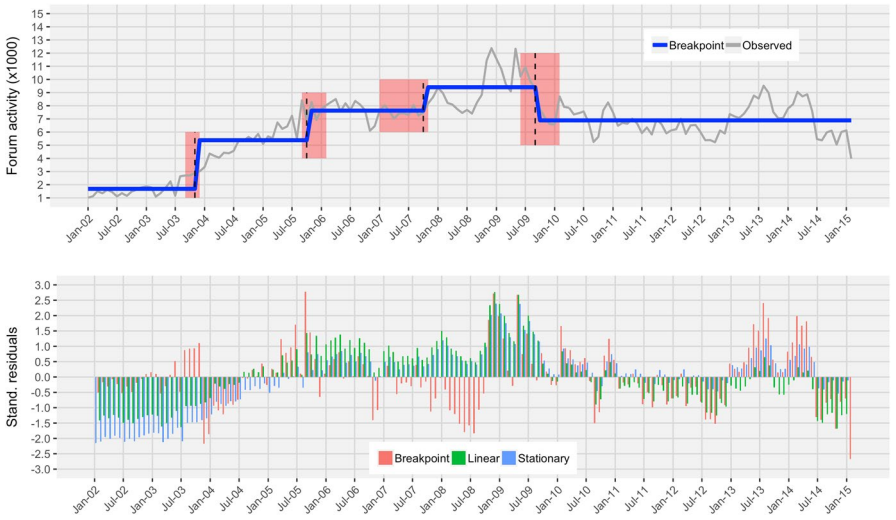

Fig. 1 Forum activity (in thousand posts) per month. Top: the observed data (grey) and the breakpoint model (blue) with structural breaks indicated by the dashed line. The red areas represent the 95% confidence interval of the breakpoints. Bottom: the standardised residuals (z scores) of the breakpoint, linear and stationary model

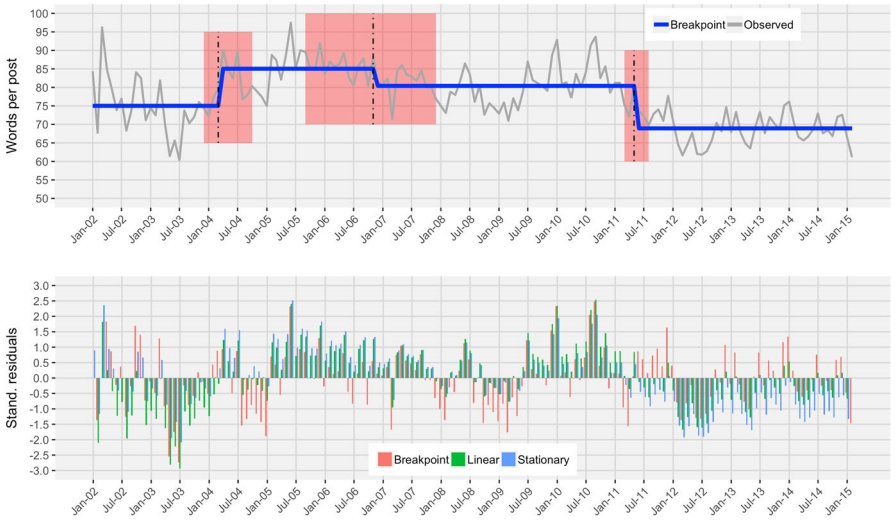

Fig. 2 Words per post per month. Top: the observed data (grey) and the breakpoint model (blue) with structural breaks indicated by the dashed line. The red areas represent the 95% confidence interval of the breakpoints. Bottom: the standardised residuals (z scores) of the breakpoint, linear and stationary model

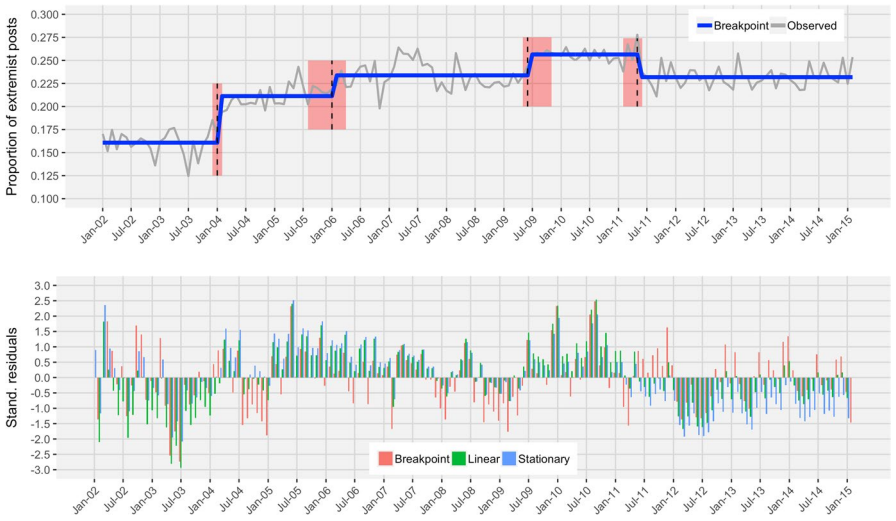

Fig. 3 The proportion of extremist posts per month. Top: the observed data (grey) and the breakpoint model (blue) with structural breaks indicated by the dashed line. The red areas represent the 95% confidence interval of the breakpoints. Bottom: the standardised residuals (z scores) of the breakpoint, linear and stationary model

Proportion of extremist posts

There was no evidence in favour of a stationary model, Dickey–Fuller statistic = − 1.79, $p = 0.663$, and fit indices (Table 2) suggests a breakpoint model with four structural regime shifts at 1/2004 (increase from 16.08 to 21.13%), 1/2006 (increase to 23.37%), 6/2009 (increase to 25.65%), and 5/2011 (decrease to 23.18%). In the forum at its peak for the 2 years between 6/2009 and 5/2011, more than one-fourth of all posts met our definition of extremist (Fig. 3). Although the last regime shift was a decrease, the overall proportion of extremist posts was still high at 23.18%. Overall, the number of potentially extremist content posts remained high compared to the forum's early years.

Intensity of extremist language

For the continuous measurement of extreme language, we cannot reject the non-stationarity null hypothesis, Dickey–Fuller statistic = − 1.91, $p = 0.616$. The breakpoint model with a regime shift at 3/2004 (increase) and 5/2011 (decrease) outperformed the stationary and linear models and indicated a plateau in the intensity of extremist language for more than 7 years between 3/2004 and 5/2011 (Fig. 4). After 5/2011, the extremist language intensity decreased to a level comparable to the forum's early period (1/2002 to 3/2004).

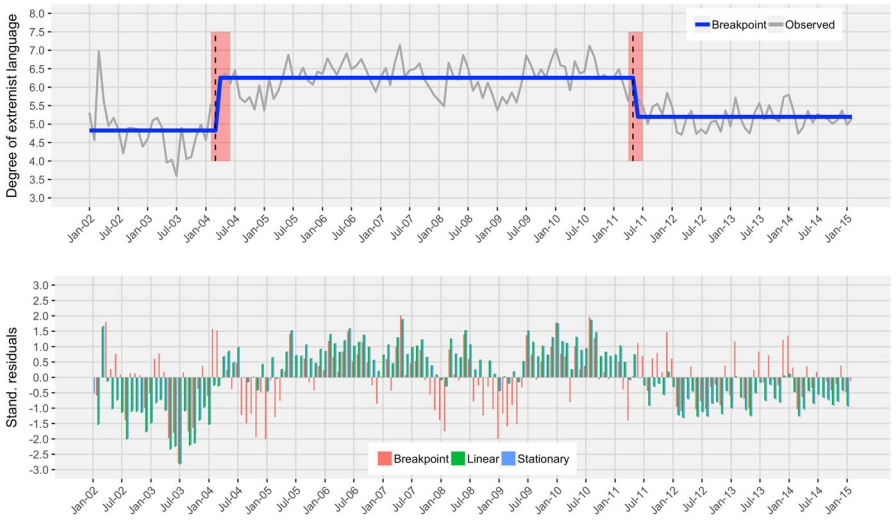

Fig. 4 The degree of extremist language over time per month. Top: the observed data (grey) and the breakpoint model (blue) with structural breaks indicated by the dashed line. The red areas represent the 95% confidence interval of the breakpoints. Bottom: the standardised residuals (z scores) of the breakpoint, linear and stationary model

User-level analysis

We assessed whether a select number of users dominated forum activity and the number of extremist posts. If there were no concentration of the overall number of posts, we would expect that the cumulative percentage of users is similar to the cumulative percentage of posts. For example, a non-concentrated pattern would indicate that 5% (20%, 60%) of the users are responsible for 5% (20%, 60%) of the posts. A measure to quantify "inequality" in the forum contributions is the Gini coefficient (see [31], see Appendix 1 for Lorenz curves). The Gini coefficient expresses the degree of inequality as a single number between 0 (perfect equality) and 1 (perfect inequality).

The Gini coefficient was 0.85 (99% CI 0.85; 0.86) for the overall number of posts and 0.90 (99% CI 0.89; 0.91) for the number of extremist posts. The most active 10% of the forum users accounted for 80.31% of the overall posts, and the most active 20% for 89.91% of all posts. The high degree of concentration on a few "super users" led us to explore whether the forum engagement of super users progressed in a different temporal manner than the forum engagement of normal users.

User segmentation: We compared the most super forum users defined as those with the number of posts in the 99th percentile (min. 398 posts, $n = 417$, accounting for 38.87% of the overall posts) with the remaining of users ($n = 41{,}077$, accounting for 61.13% of the posts).

User-level differences: To test whether the super forum users differed from the normal users on the variables reported in Table 1, we conducted independent t tests and report the sample size-standardised effect size Hedges' g with its 95% confidence intervals. Super forum users' posts were significantly longer than those of normal users ($g = -0.09$), contained more profane language ($g = -0.10$), and a less

Table 3 Means (SDs) for super users and normal users with test statistics

	Normal forum users	Super forum users	*p* value	Hedges' *g* (95% CI)
Post length	71.34 (139.67)	87.03 (169.28)	< 0.001	−0.09 [−0.19; 0.00]
Profane language (%)	5.34 (5.21)	5.87 (5.29)	< 0.001	−0.10 [−0.20; 0.00]
Racial slurs (%)	2.39 (2.94)	2.38 (2.88)	0.091	0.00 [−0.09; 0.10]
Sentiment	0.08 (0.30)	0.04 (0.29)	< 0.001	0.14 [0.04; 0.23]
Extremist posts per month (%)	21.10 (40.80)	26.16 (43.95)	< 0.001	−0.12 [−0.21; −0.02]
Extremist language per month	5.25 (10.37)	6.81 (13.76)	< 0.001	−0.11 [−0.21; −0.02]
Mean age of oldest post (days)	2407.35 (1276.23)	2961.73 (1110.90)	< 0.001	−0.43 [−0.53; −0.34]
Mean activity range (days)	273.30 (605.42)	2119.72 (1192.23)	< 0.001	−3.01 [−3.12; 2.91]

The Hedges' *g* effect size denotes small ($g < 0.20$), moderate ($g < 0.50$) and large ($g < 0.80$) effects [32]. Negative effect sizes indicate that the variable was more prevalent in super forum users than normal forum users

positive sentiment ($g = 0.14$, see Table 3). Moreover, the super users had a higher rate of extremist language posts ($g = -0.12$) and a higher intensity in extremist language ($g = -0.11$). These findings suggest that the users who engaged most with the forum also—on average—used more extreme language.

We also examined how the user levels differed on temporal variables on the forum.[2] Super users' first posts dated back significantly longer than those of the normal users. On average, the days since their first post was considerably higher in super users than in normal users ($g = -0.43$). Super users were active for a much longer period of time than normal users—almost eight times as long—measured as the number of days between their first and last post on the forum in the current dataset ($g = -3.01$).

Trajectory extraction: The individual temporal trajectories for the posts made on the forum per user were obtained by standardising the temporal progression to a scale from 1 to 100, and by applying a discrete cosine transformation to the post frequency scores [25, 26, 33].[3] The forum engagement was scaled to a score ranging from 0 (minimal forum engagement) to 1 (maximal forum engagement). This allowed us to extract one temporal progression profile for each user that is comparable with all others in a relative sense: for each user, their lowest engagement was scored as 0 and their highest engagement with 1. Figure 5 shows the averaged temporal trajectories for the users in the 99th percentile (right) and the rest (left). We fitted stationary, linear and breakpoint models to the data. For both

[2] We thank an anonymous reviewer for his/her suggestion to look at this aspect.

[3] We used a low-pass filter size of 20. Note that the low-pass filter required a sufficient number of observations per user (here: months with at least one post). Users with fewer than 24 months of posts were, therefore, excluded from the analysis.

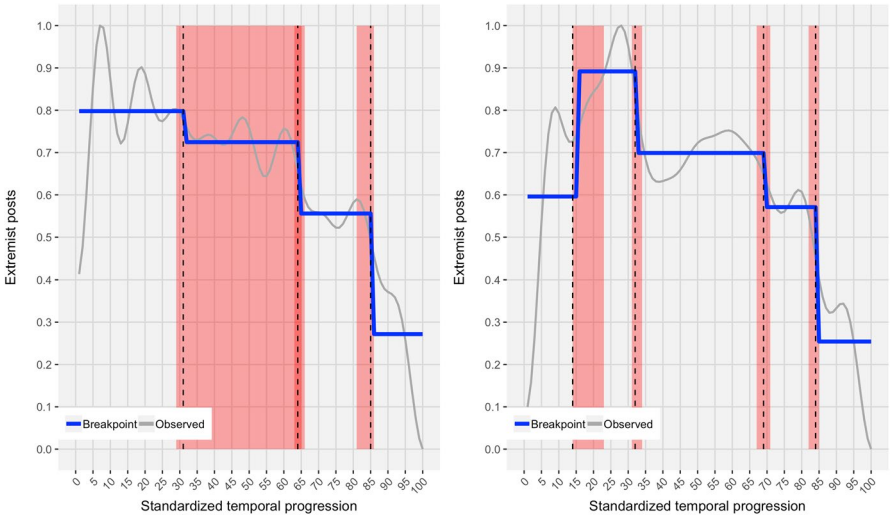

Fig. 5 Averaged temporal trajectory shapes of forum engagement (grey) and the best fitting breakpoint model (blue). Left: normal users with a post count lower than 398 (< 99th percentile). Right: super users with a post count higher than 398 (99th percentile)

user groups, the ADF test did not allow for the rejection of the stationarity null hypothesis, and the breakpoint models achieved the best model fit.[4]

The trajectory plots suggest that, in contrast to the "super users" (highest 1%), the regular users' forum engagement, on average, decreased over time with three breakpoints (for user-level trajectories of the number of extremist posts, see Appendix 2). Contrary to the forum-level analyses, the breakpoints can only be dated relative to each user's time on the forum. The breakpoints correspond to 31%, 68% and 85% of the users' temporal progression on the forum. Put differently, on average, the users started with a high forum engagement followed by two decreases roughly after one-third and two-thirds of their time on the forum. The decreases at 68% and 85% show marked drops in forum engagement. Overall, the left plot suggests that for regular users, the initial, relatively high posting frequency, quickly faded and decreased in three successive steps.

For the super users, we observe a marked increase after 15% of the users' temporal progression, which lasts until 32%. From then onwards, we observe three successive structural decreases. These findings indicate that, aside from their overall post frequency, super users differ, above all, from normal users by displaying an increase in the first third of the temporal progression of forum activity. Interestingly, both user profiles show an evident decrease in forum engagement as the user progresses on the forum. Note that the decreases of forum engagement do not necessarily imply the disengagement of the user. What the two plots in Fig. 5 suggest is that both user groups—on average—reach their lowest forum activity

[4] Normal users: Dickey–Fuller statistic = − 2.62, $p = 0.32$. Superusers: Dickey–Fuller statistic = − 0.14, $p = 0.99$.

towards the end of the range of the available dataset. While this is roughly in line with the trend for the whole forum (Fig. 1), it is possible that some users re-engaged after a decrease in activity.

Discussion

This investigation sought to understand the temporal evolution of language on a niche internet forum. Specifically, this paper tested whether potentially extremist language develops in a linear fashion or in discrete structural breaks. We examined both the engagement with the forum and the use of language on the forum. As a whole, the time-series analysis suggests that structural break models better capture temporal development than stationary or linear change models. The number of posts made on the forum increased until Sept. 2009 and then declined. Similarly, the average post length reached a peak between March 2004 and Nov. 2006 but declined as time progressed. These two engagement measures suggest that the general interest, as evidenced by actual activity on the forum, declined towards the end of the data's time range.

In contrast, the proportion of potentially extremist posts increased in three steps until its highest rate around May 2011, where more than one in four posts were considered potentially extremist. Although there was a decrease afterwards, the proportion remained high. Here, the analysis indicates that Stormfront became more extreme in its language and did so in a stepwise manner. Likewise, the intensity of potentially extremist language occurred in an increase–decrease pattern with a broad plateau of more than seven years until May 2011, after which it approached its normal level.

Taken together, our results allow for two main conclusions about the temporal evolution of the whole forum. First, the forum activity, as well as the presence of potentially extremist language, tended to develop in discrete steps rather than as a continuous linear change. For all outcome measures, a model with structural breakpoints captured the temporal dynamics of the data better than a simple linear model. Second, the findings suggest that the right-wing forum Stormfront did not become more extreme in language use as time progressed. We observe a decline on all outcome measures after early summer of 2011—none of the outcome measures displayed an increase after that period, challenging the view that the forum attracted more user activity, more user engagement, or more potentially extremist posts. This finding stands in stark contrast with previous work on right-wing extremist online spaces [18] and propaganda [34], where steadily increasing participation [18] and negative language [34] were found over time, potentially indicative of polarisation or radicalisation.

Interpreting the change points

Since breakpoint models were the best fit for both forum engagement and extremist language, this raises the question whether specific events at the time of each breakpoint gave rise to an increase or decrease in the trajectories. For instance, post-May 2011, the study observes a decrease in extreme language on Stormfront.

In that month, various major events took place, such as the killing of Osama Bin Laden [35]. However, since the current approach made use of aggregate engagement and language data, we are unable to draw conclusions on whether specific events affected changes in temporal development. Future in-depth examinations of forum content may reveal whether specific historical events attracted much attention on the forum, after which potential hypotheses can be formed about whether the event impacted forum behaviour.

Superusers versus normal users

There was clear evidence of concentration of forum activity on a few "super" users akin to the Pareto principle (20/80 rule). A mere 10% of users were responsible for more than 80% of the posts, and 20% of the users accounted for almost 90% of the posts. A comparison of the relative temporal trajectories of the individual users' post frequency between super users and regular users led to four core conclusions. First, similar to the forum-wide analysis, structural breakpoint models explained the temporal development better than stationary or linear models. Second, both user types reached their relative maximum number of posts early within the first 25% of their forum activity. This finding relates to previous work on radical posting behaviour on a Canadian right-wing forum [18], where the frequency of anti-Semitic, anti-black, and anti-LGBTQ posts tended to tail off after the first 19 months of forum activity (12% of a total of 160 months in the dataset). It must, however, be noted that posting behaviour did consider not only the number of posts, but also negative language use in the posts [18].

Third, both user types reached their relative minimum forum activity towards the end of their time on the forum (within the temporal constraints of the current data set). Potentially, the latter might be a consequence of users' disengagement from with the forum before they ceased to contribute actively altogether. In many ways, the decreasing pattern observed in this study mirrors research on disengagement from real-world extremist settings. It is a gradual process punctuated by different key catalytic events [36]. Future analyses might identify whether similar push (such as disagreement with violence, disillusionment, and/or fear of confinement) and pull factors (such as changes in personal circumstances, social relationships, and competing obligations) similarly spark disengagement from virtual extremist settings [37].

Fourth, normal users started with a structural regime at their peak of post frequency and then disengaged in three successive regime decreases. Super users, on the other hand, displayed a marked increase of almost 30 percentage points after 15% of their time on the forum. That is, their posting behaviour changed to a peak activity level after an initial, more moderate posting frequency. A possible explanation of that pattern is that the super users initially explored the forum more passively and then rapidly started posting at a high rate. However, the peak posting frequency lasted only briefly and was followed by two moderate decreases and one marked decrease at around 85% of their time on the forum. Possibly the peak frequency was challenging to maintain, faded off to a lesser albeit still relatively high rate and then wholly dissipated before the user disengaged from active forum participation.

Statistical testing revealed that super users wrote longer posts, resorted to more profanity, used more negative sentiment and had a higher rate and intensity of extremist language. Albeit with a small absolute effect size, these findings hint at an overlap between those who engage extensively with the forum by posting often and those who account for the majority of extremist language. Indeed, 85.13% of the super users were also in the 99th percentile of the users with the most extremist posts. Here, it would be interesting to replicate findings from Shrestha et al. [14] on the identification of radical users by the use of jargon and highly contextualised language on a right-wing forum. Finally, the super users were active contributors to the forum well before the normal users and had a forum activity span of almost eight times that of the normal users. Further examinations could shed light on whether the forum user age plays a role in the use and adoption of extremist language.

Limitations and future directions

While the current investigation provided a first look and quantitative analysis at the temporal dynamics of an online forum, several limitations are worth discussing.

First, the sample studied here is only one of many niche internet forums (for others, see [5, 14, 18]). Thus, absent any cross-forum examination of the temporal dynamics, our findings cannot be generalised to other forums or even other right-wing forums. Future research could look at whether user engagement and the use of language follow similar patterns across platforms and political orientation. Similarly, the findings reported here should be strictly interpreted within the constraints of the dataset and its temporal range (ending in Feb. 2015). Since the end of the dataset, more temporal shifts might have happened and more users could have disengaged from the forum and new members could have started with actively participating in the forum. For example, since 2015, some users who in our analysis disengaged, could have re-engaged, and some users of are not identified as disengagers until 2015 could have done so in the succeeding years.

Second, as with most attempts measuring linguistic constructs, our operationalization of "potentially extremist language" is just one of many possible operationalizations. Ideally, as a research community, agreement about the operationalizations of constructs relevant to the study of extremist language can be found (e.g., similar to the LIWC software used in psycholinguistics). Equally, a challenge for related research is the potential language adaptation of the forum users. Some studies suggest that some Stormfront users deliberately choose a moderate racist terminology to resonate with wider audiences [38], while others suggest a gatekeeper function of Stormfront forum moderators (i.e. in preventing access of those who might threaten the unity of the forum, [39]). Interestingly, such a language adaptation could possibly underlie the broad plateau that we find for the intensity of potentially extremist language: possibly, the forum managed to contain the language in a manner that allowed the website to stay online despite the often racial and nationalist tone.

Third, we did not examine the content of the posts in-depth. Especially for vague concepts such as extremist language, the actual content might sketch a more nuanced picture of the forum dynamics. A possible consequence is that our rule-based

operationalization of extreme language did not capture extremist content that we might have been able to find with a more content-specific approach. Future work could, therefore, look at the actual content (e.g. frequently used *n*grams) and the development of topics over the course of the forum to add a dimension of analysis that could both assess the usefulness of the operationalization and offer insights into the content of the posts (e.g. what are the users talking about and how). Moreover, we excluded posts shorter than 15 words to avoid that short posts or single-phrase comments to avoid that our operationalization of extremist language is inflated by short expressions of extremist language and to ensure such an inflation is not perpetuated to the trajectory extraction (the analyses in Appendix 3 show that the findings do not change if the whole dataset is used).

Fourth, the user-level analysis presented here could be expanded to anticipate the trajectories of users (or the whole forum). An avenue for future research could be to explore how and whether the progression through a forum's life span can be forecasted and whether phase transitions could be predicted. The discrete nature of the temporal sequence suggested by our analysis could further allow for forecasting not only of the progression but also of the occurrence of critical phase transitions to more extreme regimes. Ultimately, this could help in mitigating the trade-off between allowing free speech and preventing hate speech online because the passing of critical thresholds could be predicted. This would allow policymakers to make transparent, evidence-based decisions concerning risk management on such online spaces.

Conclusions

The analysis of forum engagement and the presence of extremist language revealed that a discrete, stepwise pattern more likely underpins the temporal evolution of the right-wing forum Stormfront. Within the temporal constraints of the data available, this work found that the forum rhetoric did not become more aggressive over time. With future work on the quantitative analysis of forum evolution, potential escalations could be anticipated and mitigation efforts could be formulated.

Acknowledgements This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 758834).

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix 1

See Fig. 6.

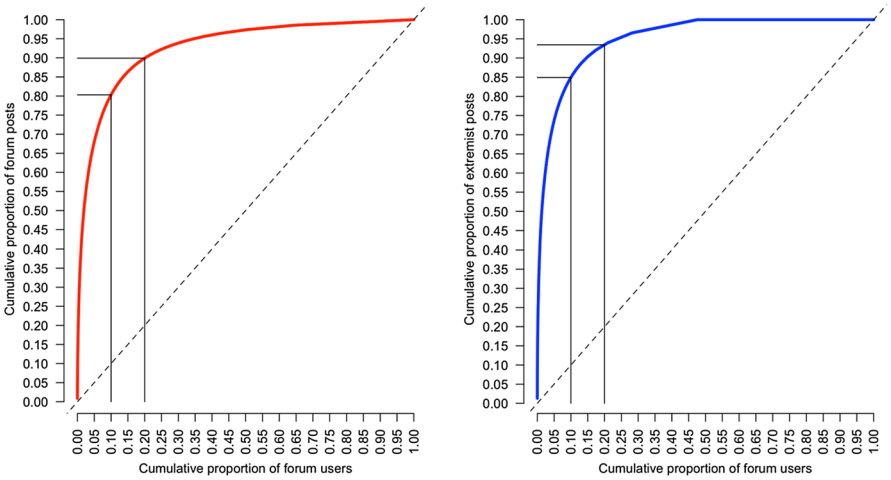

Fig. 6 Lorenz curve for the number of overall posts on the forum (left) and the number of extremist posts (right)

Appendix 2

See Fig. 7.

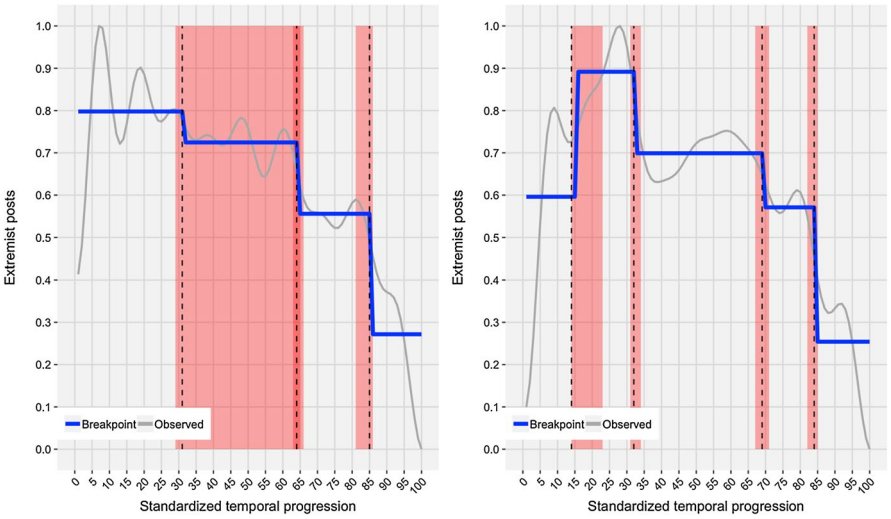

Fig. 7 Averaged temporal trajectory shapes of extremist posts. Left: users with an extremist post count lower than 93 (< 99th percentile). Right: users with a post count higher than 93 (99th percentile)

Appendix 3

The decision to remove posts shorter than 15 words after stop word removal was made to avoid the inclusion of too short posts that add little to the phenomenon under study.

Since the effects of that decision on the outcomes are not known, we re-ran the core analyses of the current paper on the full dataset. We removed all posts that were shorter than one word (i.e. either were empty or did not contain at least one non-stop word, $n = 15{,}187$). That sample consisted of 1,654,513 posts. Table 4

Table 4 Descriptive statistics for all posts with at least one word

	Mean	SD	Median	Range
Post length	50.41	123.69	22.00	1; 10,388
Profane language (%)	5.37	7.72	3.08	0.00; 100.00
Racial slurs (%)	2.52	5.19	0.00	0.00; 100.00
Sentiment	0.08	0.37	0.04	−2.25; 2.53
Number of posts per month	10,451	4215	11,216	1473; 20,916
Post length per month	50.41	123.69	22.00	1; 10,338
Extremist posts per month (%)	15.05	2.12	15.22	7.99; 19.24
Extremist language per month	3.72	0.53	3.72	2.18; 5.17

Table 5 Model fit indices of the stationary, linear and breakpoint model for each of the four forum-level outcome measures

Outcome measure	Model	Evaluation metric			
		AIC	BIC	MAE	RMSE
Forum activity (number of posts)	Stationary model	3088.85	3094.98	3207.95	4201.70
	Linear model	3029.02	3038.39	2821.48	3456.99
	Breakpoint model (5)	**2814.08**	**2835.25**	**1278.11**	**1706.29**
Number of words per post	Stationary model	1045.71	1051.84	5.37	6.54
	Linear model	994.55	1003.74	4.38	5.53
	Breakpoint model (4)	**917.95**	**936.32**	**3.24**	**4.25**
Proportion posts with extremist language	Stationary model	−766.05	−759.92	0.0260	0.0212
	Linear model	−834.57	−825.38	0.0139	0.0169
	Breakpoint model (4)	**−1013.08**	**−994.71**	**0.0073**	**0.0094**
Average extremist language score per post	Stationary model	253.29	259.42	0.4483	0.5326
	Linear model	252.95	262.14	0.4357	0.5286
	Breakpoint model (3)	**115.50**	**130.81**	**0.2510**	**0.3379**

The integer in brackets after the breakpoint model indicates the number of regime shifts in the model. The best model fit is highlighted in bold

AIC Akaike's information criterion, *BIC* Bayesian information criterion, *MEA* mean absolute error, *RMSE* root mean square error

shows that the descriptive statistics are in line with those of the sample after removal of short posts (except for the variables related to the absolute length of the posts).

Table 5 shows that the key conclusions about the nature of the evolution of the forum engagement and the use of language are consistent with the analyses run on the sample used in the main analysis.

References

1. Burke, J. (2019). Norway mosque attack suspect "inspired by Christchurch and El Paso shootings. *The Guardian*. Retrieved from https://www.theguardian.com/world/2019/aug/11/norway-mosque-attack-suspect-may-have-been-inspired-by-christchurch-and-el-paso-shootings.
2. Hern, A. (2017). Stormfront: "Murder capital of internet" pulled offline after civil rights action. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/aug/29/stormfront-neo-nazi-hate-site-murder-internet-pulled-offline-web-com-civil-rights-action.
3. Southern Poverty and Law Center. (2014). *White Homicide Worldwide*. Retrieved from https://www.splcenter.org/20140331/white-homicide-worldwide.
4. Wong, J. C. (2019). 8chan: The far-right website linked to the rise in hate crimes. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website.
5. Hine, G., Onaolapo, J., Cristofaro, E. D., Kourtellis, N., Leontiadis, I., Samaras, R., Blackburn, J., et al. (2017). Kek, Cucks, and God Emperor Trump: A measurement study of 4chan's politically incorrect forum and its effects on the web. In *Proceedings of the eleventh international AAAI conference on web and social media* (p. 10).
6. Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., & Blackburn, J. (2018). What is gab: A bastion of free speech or an alt-right echo chamber. In *Companion proceedings of the web conference 2018* (pp. 1007–1014). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3184558.3191531.
7. Iqbal, N., & Townsend, M. (2019). Christchurch mosque killer's theories seeping into mainstream, report warns. *The Observer*. Retrieved from https://www.theguardian.com/world/2019/jul/07/christchurch-mosque-killer-ideas-mainstream-social-media.
8. Levin, S. (2017). Tech giants team up to fight extremism following cries that they allow terrorism. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/jun/26/google-facebook-counter-terrorism-online-extremism.
9. Roy, E. A. (2019). Christchurch call: Details emerge of Ardern's plan to tackle online extremism. *The Guardian*. Retrieved from https://www.theguardian.com/world/2019/may/13/christchurch-call-details-emerge-of-arderns-plan-to-tackle-online-extremism.
10. Siegel, A. A., Nikitin, E., Barbera, P., Sterling, J., Pullen, B., Bonneau, R., Tucker, J. A., et al. (2018). Measuring the prevalence of online hate speech, with an application to the 2016 U.S. election, 22.
11. van der Vegt, I., Mozes, M., Gill, P., & Kleinberg, B. (2019). Online influence, offline violence: Linguistic responses to the "Unite the Right" rally. arXiv:1908.11599 [cs]. Retrieved from http://arxiv.org/abs/1908.11599.
12. van der Vegt, I., Gill, P., Macdonald, S., & Kleinberg, B. (2019). Shedding light on terrorist and extremist content removal. *Global Research Network on Terrorism and Technology*. Retrieved from https://rusi.org/sites/default/files/20190703_grntt_paper_3.pdf.
13. Kaati, L., Shrestha, A., & Sardella, T. (2016). Identifying warning behaviors of violent lone offenders in written communication. In *2016 IEEE 16th international conference on data mining workshops (ICDMW)* (pp. 1053–1060). https://doi.org/10.1109/ICDMW.2016.0152.
14. Shrestha, A., Kaati, L., & Cohen, K. (2017). A machine learning approach towards detecting extreme adopters in digital communities. In *2017 28th international workshop on database and expert systems applications (DEXA)* (pp. 1–5). https://doi.org/10.1109/DEXA.2017.17.

15. Abbasi, A., & Chen, H. (2007). Affect intensity analysis of dark web forums. In *2007 IEEE intelligence and security informatics* (pp. 282–288). New Brunswick: IEEE. https://doi.org/10.1109/ISI.2007.379486.
16. Figea, L., Kaati, L., & Scrivens, R. (2016). *Measuring online affects in a white supremacy forum* (pp. 85–90). IEEE. https://doi.org/10.1109/ISI.2016.7745448.
17. Scrivens, R., Davies, G., & Frank, R. (2018). Searching for signs of extremism on the web: An introduction to sentiment-based identification of radical authors. *Behavioral Sciences of Terrorism and Political Aggression, 10*(1), 39–59. https://doi.org/10.1080/19434472.2016.1276612.
18. Scrivens, R., Davies, G., & Frank, R. (2018). Measuring the evolution of radical right-wing posting behaviors online. *Deviant Behavior*. https://doi.org/10.1080/01639625.2018.1556994.
19. Park, A. J., Beck, B., Fletche, D., Lam, P., & Tsang, H. H. (2016). Temporal analysis of radical dark web forum users. In *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)* (pp. 880–883). https://doi.org/10.1109/ASONAM.2016.7752341.
20. Macnair, L., & Frank, R. (2018). Changes and stabilities in the language of Islamic state magazines: A sentiment analysis. *Dynamics of Asymmetric Conflict, 11*(2), 109–120. https://doi.org/10.1080/17467586.2018.1470660.
21. Rowe, M., & Saif, H. (2016). Mining pro-ISIS radicalisation signals from social media users. In *ICWSM 2016* (p. 10).
22. Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2019). Auditing radicalization pathways on YouTube. arXiv:1908.08313 [cs]. Retrieved from http://arxiv.org/abs/1908.08313.
23. Rinker, T. (2018). *lexicon: Lexicon data*. Retrieved from http://github.com/trinker/lexicon.
24. Wiegand, M., Ruppenhofer, J., Schmidt, A., & Greenberg, C. (2018). Inducing a lexicon of abusive words—A feature-based approach. In *Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies, Volume 1 (Long Papers)* (pp. 1046–1056). New Orleans, Louisiana: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1095.
25. Kleinberg, B., Mozes, M., & Van der Vegt, I. (2018). Identifying the sentiment styles of YouTube's vloggers. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3581–3590).
26. Soldner, F., Ho, J. C., Makhortykh, M., Van der Vegt, I., Mozes, M., & Kleinberg, B. (2019). Uphill from here: Sentiment patterns in videos from left- and right-wing YouTube news channels. In *Workshop on natural language processing and computational social science*, *NAACL*.
27. Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica, 64*(4), 813. https://doi.org/10.2307/2171846.
28. Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v007.i02.
29. Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis, 44*(1–2), 109–123. https://doi.org/10.1016/S0167-9473(03)00030-6.
30. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.
31. Delbosc, A., & Currie, G. (2011). Using Lorenz curves to assess public transport equity. *Journal of Transport Geography, 19*(6), 1252–1259. https://doi.org/10.1016/j.jtrangeo.2011.02.008.
32. Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.
33. Jockers, M. (2015). *Revealing Sentiment and Plot Arcs with the Syuzhet Package*. Retrieved from http://www.matthewjockers.net/2015/02/02/syuzhet/.
34. Vergani, M., & Bliuc, A.-M. (2015). The evolution of the ISIS' language: A quantitative analysis of the language of the first year of Dabiq magazine. *Security, Terrorism and Society, 2,* 7–20.
35. Walsh, D., Adams, R., & MacAskill, E. (2011). Osama bin Laden is dead, Obama announces. *The Guardian*. Retrieved from https://www.theguardian.com/world/2011/may/02/osama-bin-laden-dead-obama.
36. Horgan, J. (2008). Individual disengagement: A psychological analysis. In T. Bjorgo & J. Horgan (Eds.), *Leaving terrorism behind: Individual and collective disengagement* (pp. 35–47). https://doi.org/10.4324/9780203884751-10.

37. Windisch, S., Simi, P., Ligon, G. S., & McNeel, H. (2016). Disengagement from ideologically-based and violent organizations: A systematic review of the literature. *Journal for Deradicalization, 9,* 1–38.
38. Meddaugh, P. M., & Kay, J. (2009). Hate speech or "reasonable racism?" The other in Stormfront. *Journal of Mass Media Ethics, 24*(4), 251–268. https://doi.org/10.1080/08900520903320936.
39. De Koster, W., & Houtman, D. (2008). "Stormfront is like a second home to me": On virtual community formation by right-wing extremists. *Information, Communication & Society, 11*(8), 1155–1176. https://doi.org/10.1080/13691180802266665.